様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　６　月　１９　日現在

機関番号：　１２１０２
研究種目：　基盤研究(C)（一般）
研究期間：　2015～2017
課題番号：　１５Ｋ００４３４

研究課題名（和文）異なり数計測アルゴリズムを多分野に適応するための最適パラメータの設定手法の開発

研究課題名（英文）Development of a method for finding a appropriate values of parameters for the cardinality count analysis applying various cases

研究代表者

佐藤　聡（Sato, Akira）

筑波大学・システム情報系・准教授

研究者番号：９０２８５４２９

交付決定額（研究期間全体）：（直接経費）　　3,600,000円

研究成果の概要（和文）：異なり数解析は，ネットワークの不適切な利用の発見手法として有益である．様々な事例に異なり数解析を適用することにより，異なり数計測アルゴリズムにおいて設定すべきパラメータの一つである閾値に対する適切な値の調査を行なった．その結果より，閾値の決定は，異なり数解析の結果の利用方法や，適応する属性値の取りうる値域など，様々なことを考慮する必要があることがわかった．
また，適用する属性値が連続値である場合には，アルゴリズムにより求めた部分集合の各要素の属性値の分散と平均が，異なり数と同様の特徴を示すことがわかった．

研究成果の概要（英文）：A cardinality count analysis is useful as a method for detecting abnormal use of network systems. We study the the appropriate value of a threshold, which is one of parameters to be set in the cardinality count algorithm, by applying the cardinality count analysis for various cases. From results of this study, we found the value of threshold must be determined in consideration of the method of utilizing the analysis result and a domain of attribute value. In addition, when the attribute value is a continuous value, we found the variance and the average of the attribute values in each element of the subset aggregated by the algorithm is shown same characteristics as the cardinality count.

研究分野：　情報工学

キーワード：　情報システム　情報通信工学　セキュア・ネットワーク

１．研究開始当初の背景

近年，インターネットは重要な社会インフラとして必要不可欠なものになってきた．そのために，より安全で安心して使えるようにするための研究が盛んに行われている．その中で，異なり数計測技術は，様々なレイヤーにおけるネットワークの不適切な利用の発見手法への適用可能性があることがわかってきた．これは，ネットワーク通信に関するデータの多くは冪即に支配されており，その特性を使えば，効率よい解析が可能であるという統計的法則と，不正行為も経済的目的を持つときにはその行動が冪則に支配され，検知が容易になるという事実に基づいている．

異なり数計測技術においては，閾値というパラメータを適切に設定する必要がある．また，異なり数計測においては，計測対象となる観測データが連続値である場合，何らかの方法で離散値に変更する必要がある．これら，閾値の設定，離散値への変更方法等は，それぞれの応用分野において，適切な方法を選択する必要がある．

２．研究の目的

先行研究では，異なり数計測アルゴリズムを用いて，１）組織内ネットワークと外部のネットワークとの境界を流れるパケットを対象とした解析，２）境界に設置したファイアウォールログを対象とした解析に用いて，ネットワーク利用動向の分析を行った．この時，アルゴリズムの動作を定めるパラメータはそれぞれの解析では異なっていることが分かった．

この計測アルゴリズムは，機器に組み込むなどの様々な応用に適用可能であるが，各々の応用分野においてパラメータをどのように定めるとよいかという手法は確立していない．本研究では，最適なパラメータ設定を効率よく定める方法の確立を目標とする．

３．研究の方法

いくつかの実データに対してアルゴリズムを適応してパラメータの最適な値について検討を行う．

・ハニーポットへ到達するパケット
・スイッチを通過するパケットをサンプリングしたパケット
・メールサーバのログのアクセス元 IP アドレスをアクセス元が所属する AS(Autonomous System）の番号に変換したもの
・筑波大学内のファイアウォールのログ
・筑波大学内の DNS フルリゾルバのログ

また，異なり数を数える属性については，連続値をとるものは直接取り扱うことができないため，統計量である，平均，分散を用いる方法を開発し，ファイアウォールログの通信量を用いて解析を行った．

４．研究成果

それぞれのデータについて異なり数解析を適用して得られた知見は以下の通りである．

１）　ハニーポットへ到達するパケットへの適用

ハニーポットへ到達するパケットに対して異なり数解析を適用した．既存研究である，ルータ・スイッチを通過するパケットを対象とした方法を適用し，ハニーポットへ通信をするホストの分別に，異なり数解析の結果を適用した．実環境において，検証実験を行い，不適切なホストの発見を行うことができた．この場合，ハニーポットへ到達するパケット数は，既存研究が対象としたパケット数よりも小さいため，閾値を大きくすること可能となる．大きくすることにより，不適切な通信が発見できるまでの時間がかかることとなる．したがって，適切な値を定めるためには，その環境や，要求要件によって異なることがわかった．

２）　スイッチを通過するパケットをサンプリングしたパケットに対する適用

既存研究として，スイッチを通過するパケット軍に対して異なり数解析を適用して，不適切な通信の発見ができることが示されている．これはサンプリングされていないパケットを対象としている．サンプリングしたパケット群に対して同様の解析を行なった．全パケット数に対する不適切な通信を行なっているパケット数の割合が，サンプリングレートよりも十分に大きい場合には，同様に検知できることがわかった．また，サンプリングレートと閾値との関係についても調査を行なったが，取り扱う属性の値域の範囲内で閾値を設定すべきこと，及び，小さい閾値では，異なり数解析の特徴を活かせないことがわかった．これについては引き続き研究を進めていく必要がある．

３）　筑波大学内のWebメールサーバのログに対する適用

メールサーバへの接続元の IP アドレスを AS 番号に変換し，ユーザ ID と AS 番号の組を対象にして異なり数解析を適用した．この事例では，不適切な利用を発見するために，どれくらいの時間にどれくらいの数のアクセスが集中していたかということが重要であることがわかった．これは異なり数解析において，アイテムセットが閾値個集まるまでの時刻を取り扱う必要があることを意味している．これについては引き続き研究を進めていく必要がある．

4） ファイアウォールのログに対する適用

一部のファイアウォール製品では，通信ログは，セッション単位で出力され，かつ，その中にそのセッションでの通信量の情報が含まれている．特定の通信（例：宛先を限定する）においては，セッション毎の通信量の変化量によって，不適切な通信であることを判別できる．そこで，異なり数解析手法を，通信量のような連続値となる属性値に対して，統計量である平均，分散を計算するように拡張を施した．そして，その拡張方式をファイアウォールのログに適用して，有効性の検証を行った．これにより，連続値となる属性については，離散化することなく異なり数分析が行えるようになった．

5） DNS フルリゾルバのログに対する適用

DNS フルリゾルバのクライアントが，その問い合わせのログを調査することにより，クライアント端末であるか，インターネットにサービスを提供するサーバであるかを識別する方法に，異なる数解析が使えるかを検討した．事前調査の結果では，クライアントとサーバでは，単位時間あたりの問い合わせ先のドメイン名の種類数に違いがあることがわかった．今後は異なり数計測の方法をそのまま適用して，識別できるかについての研究を進めていく．

いずれのデータ群に対しても，異なり数解析を行う際の閾値については，様々な値に設定し，それらの試行実験により，適切な閾値を求めることができた．しかしながら，解析結果の利用方法，対象となるデータの値域（取りうる値の範囲の大きさ）などにより，最適な閾値の値は異なることがわかった．すなわち，解析対象とるデータを分析しただけでは，適切な閾値を定めることは非常に難しいことがわかった．

一方，連続値を取り扱う方法については，いろいろな分野に応用できることが確認できた．これにより，属性値をいくつかの区間に分けて離散値にすることを行わずに解析できることがわかった．

５．主な発表論文等

（研究代表者，研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　0 件）

〔学会発表〕（計　５件）

1） 山門彩，ダークネット宛通信分析によるネットワーク管理者支援，情報処理学会　IOT 研究会，2018 年 3 月 6 日，鬼怒川温泉ホテル（栃木県）．

2） 佐藤聡，Log analysis for hijacked user account's login detection method based on an autonomous system number of access source，Workshop on Internet Architecture and Applications，2017 年 11 月 16 日，バンコク（タイ）．

3） 山門彩，ダークネット宛通信分析によるネットワーク管理者支援システム，情報処理学会　IOT 研究会，2017 年 5 月 26 日，高知工科大学永国寺キャンパス（高知県）．

4） 伊藤昂平，アクセススイッチにて取得するサンプリングパケットへの異なり数分析の適用，情報処理学会　IOT 研究会，2017 年 5 月 26 日，高知工科大学永国寺キャンパス（高知県）．

5） 渡部耕大，DNS クエリの分析によるサーバと非サーバの識別，情報処理学会　IOT 研究会，2017 年 5 月 26 日，高知工科大学永国寺キャンパス（高知県）．

〔図書〕（計　0 件）

〔産業財産権〕

○出願状況（計　０件）

○取得状況（計　０件）

〔その他〕

特になし

６．研究組織

(1)研究代表者

佐藤聡（Akira SATO）

筑波大学・システム情報系・准教授

研究者番号：９０２８５４２９

(2)研究分担者

なし

(3)連携研究者
なし

(4)研究協力者
山門彩（Aya YAMAKADO）
筑波大学・システム情報工学研究科・コンピュータサイエンス専攻（博士課程前期）
渡部耕大（Kodai WATANABE）
筑波大学・システム情報工学研究科・コンピュータサイエンス専攻（博士課程前期）
伊藤昂平（Kohei ITOH）
筑波大学・システム情報工学研究科・コンピュータサイエンス専攻（博士課程前期）